

Understanding Computer Vision

2018.02.28

Mr. Calvenn Tsuu

What is the most important sense

- Hearing
- Smell
- Touch
- Taste
- Sight

What is the most important sense

We perceive up to 80% of all impressions by means of our sight.



Processing Speed

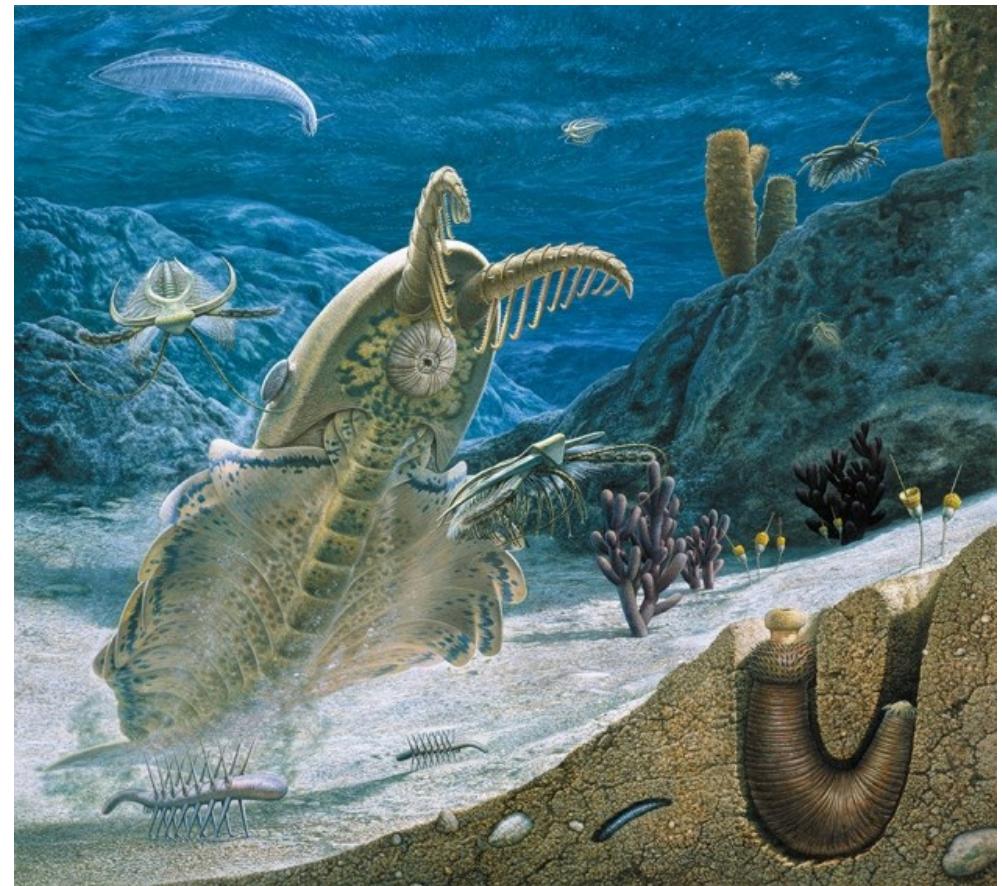
Visual information can be processed 60,000 times faster than text, and it is easier to remember.



Visual Content vs. Text Content -
Epic Faceoff with Obvious Winner

Cambrain Explosion

- 500 Million years ago
- Vision is one of the key factor
- Find more food
- Keep them alive
- In short few million years period, total number of animals exploded.



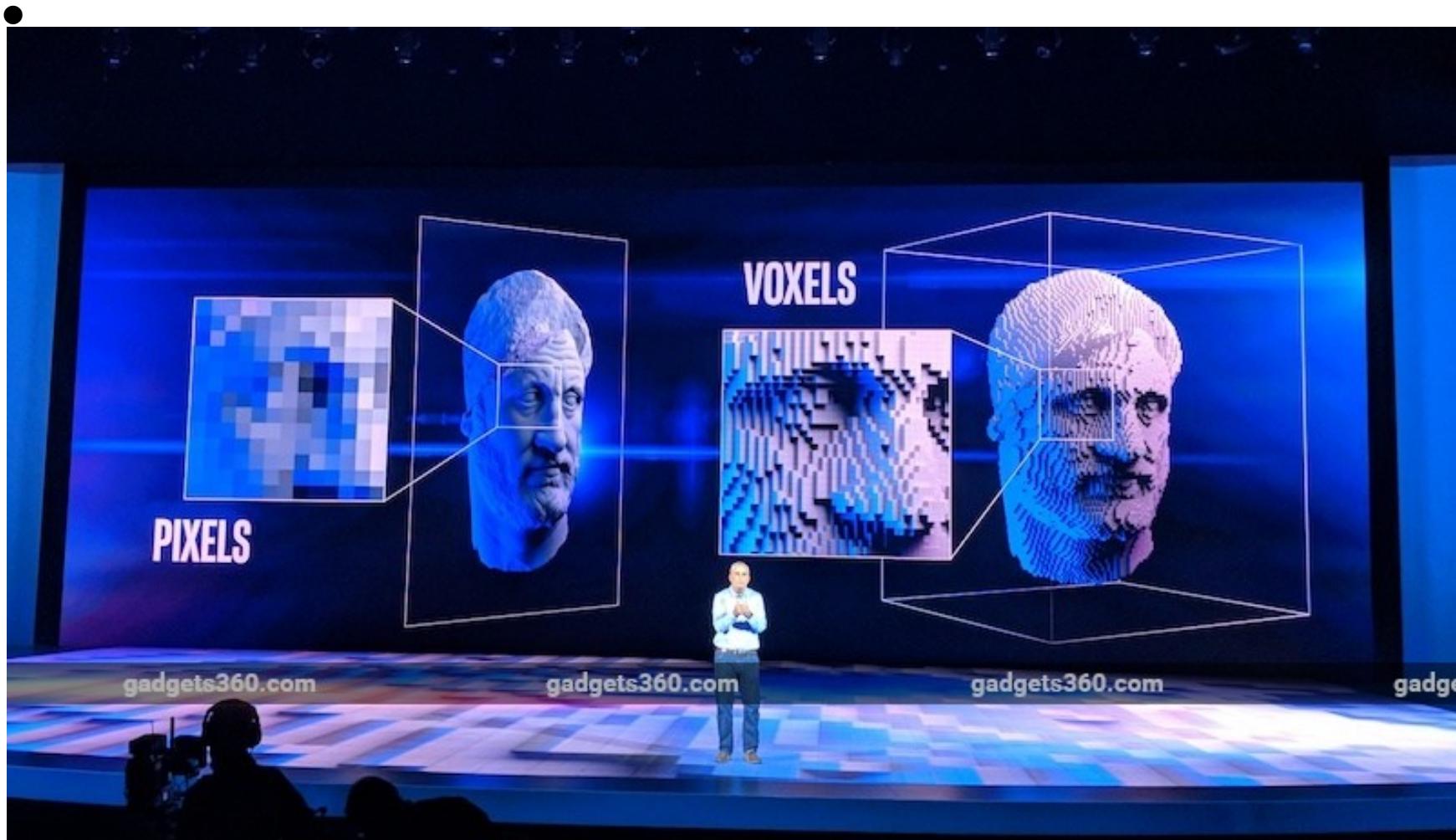
Computer Era

- 16 years ago
- Camera phone captures 640x480 pixels image
- 0.3 megapixels to 15 megapixels
- Tiny screen to Retina screen



Electronic Device Explosion

- CES 2018 Las Vegas



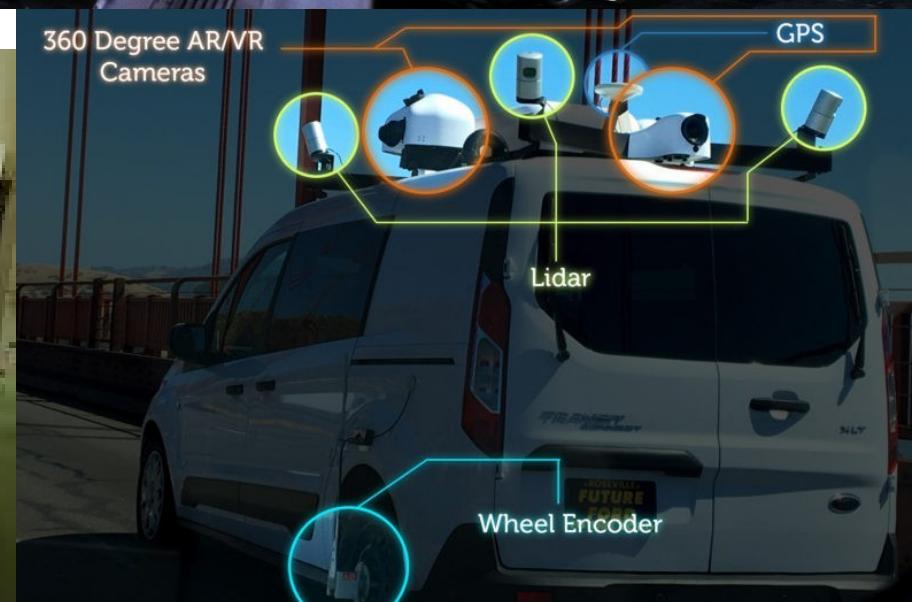
Photographer's eye vs Camera

By David Chancellor, National Geographic



Computer Vision

- Surveillance Camera
- Inspection Camera
- Self-driving Car
- Dash Camera
- Speed Camera



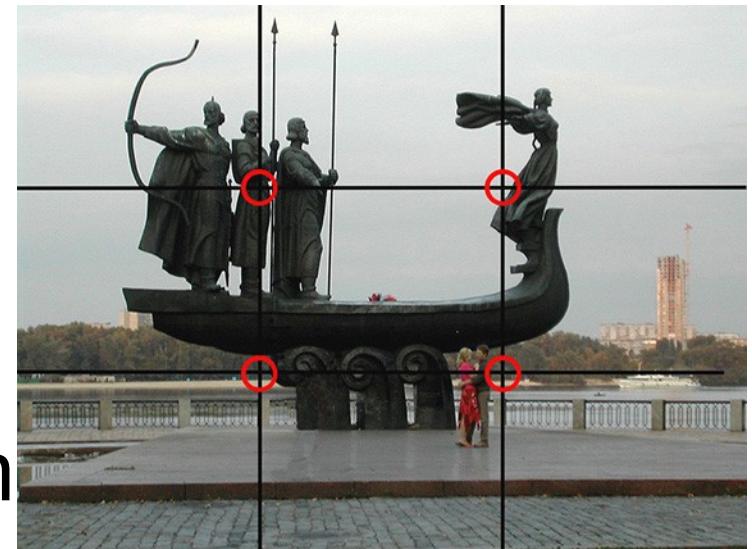
“A Critical Moment”

- A moment of decision
- Tipping Point



Photographer vs Computer Vision

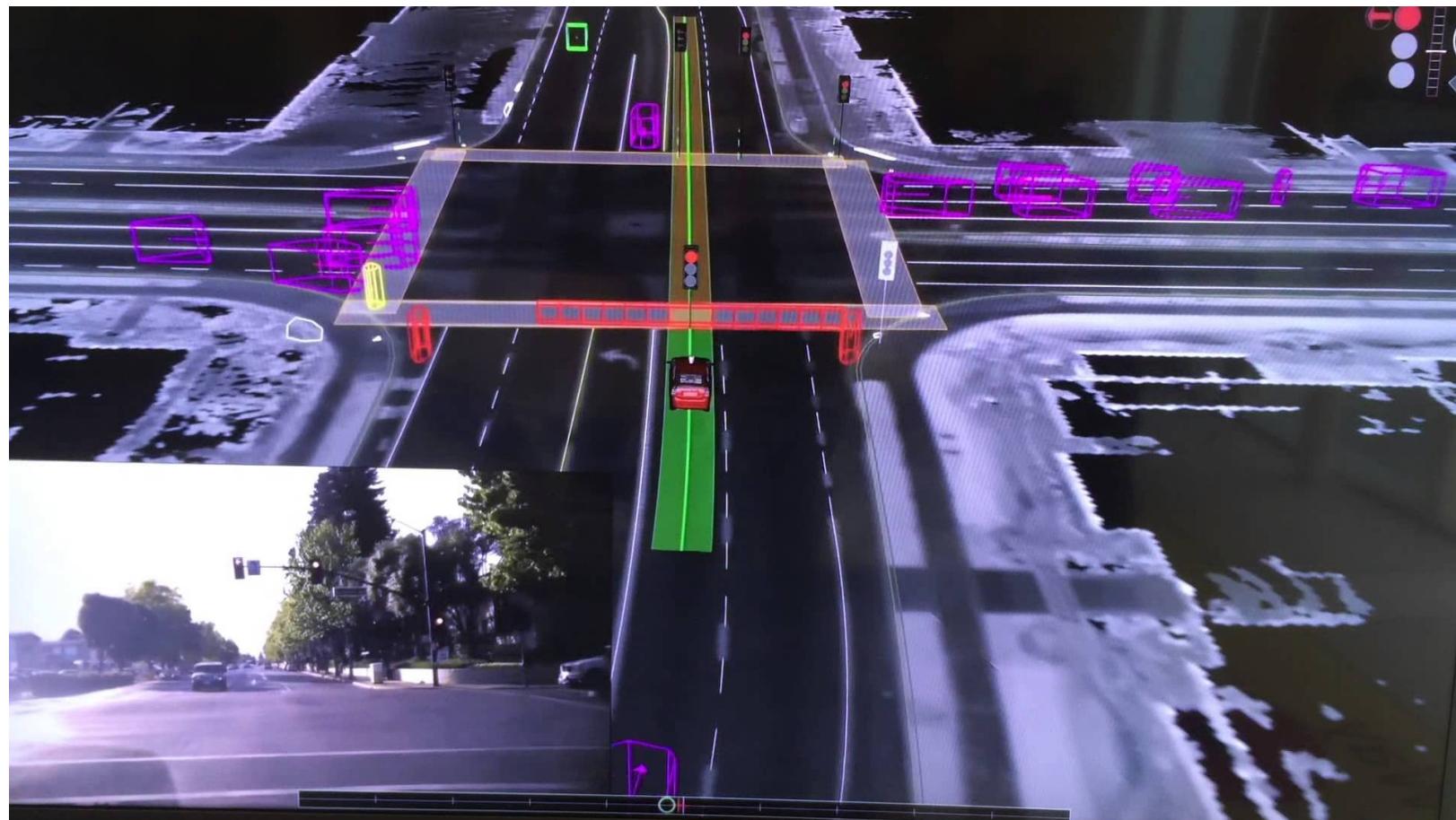
- Real photographer
- Think before photo taking
- Predict the action
- Critical Moment
- Different Angle and Position
-
- Computer Vision
- Continuous shooting
- Monitoring
- Fixed position



Computer Vision Monitoring

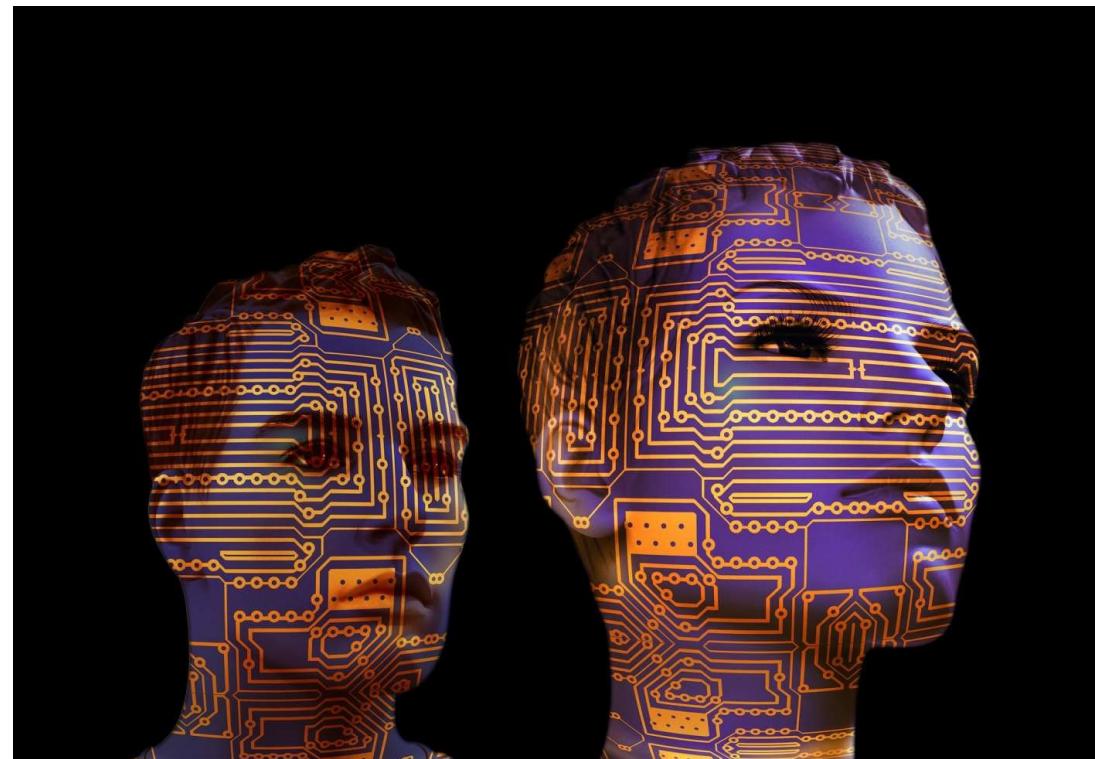
“Find footage of interest using motion based search identity activity patterns with heat mapping.”

-Shaw Ads.



Understanding Computer Vision

- The lens – How the image formed
- The sensor – How the image converted
- The raw data – How the image stored
- Aperture
- Shutter
- Post processing
- - What you can
- get from the
- raw data



Vision is based on Lighting

- Direction
- Intensity
- Reflection
- Transparency

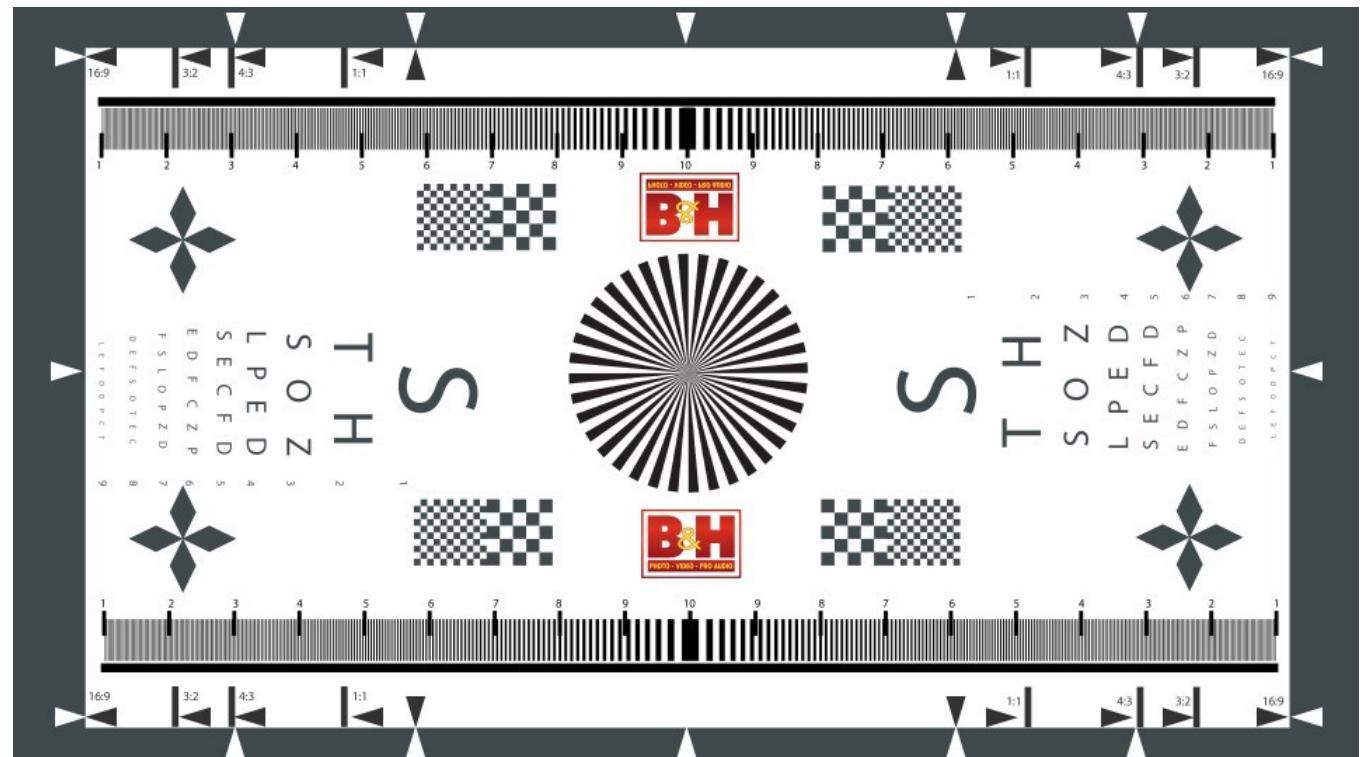


Lighting Intensity/Brightness

- 0.0001 Lux Moonless, overcast night sky
- 0.05-0.3 Lux Full Moon on a clear night
- 20-50 Lux Street Light
- 100 Lux Very dark overcast day
- 250-500 Lux Office work
- 1000 Lux Drawing/detail work
- 10000 Lux Not direct sunlight
- 100000 Lux Direct sun
- “Human can easily see from star light to sun light range translate to 30 stops.
- Digital camera is about $2^{10} = 1024:1$. ”
-
-

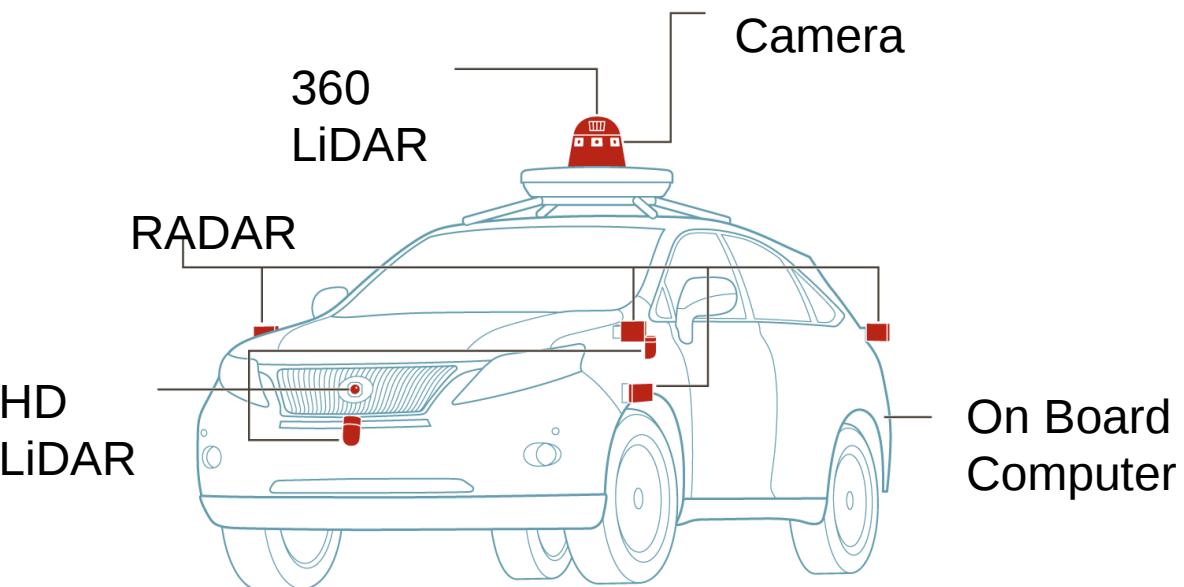
Common Camera Problems

- Near fish eye lens
 - Distortion
- Storage
 - File size
- Dynamic
- Range
- 8~10 stops
- Processing
 - Power
- Lens Resolution
 - HD 1920 x 1080 pixels
 -



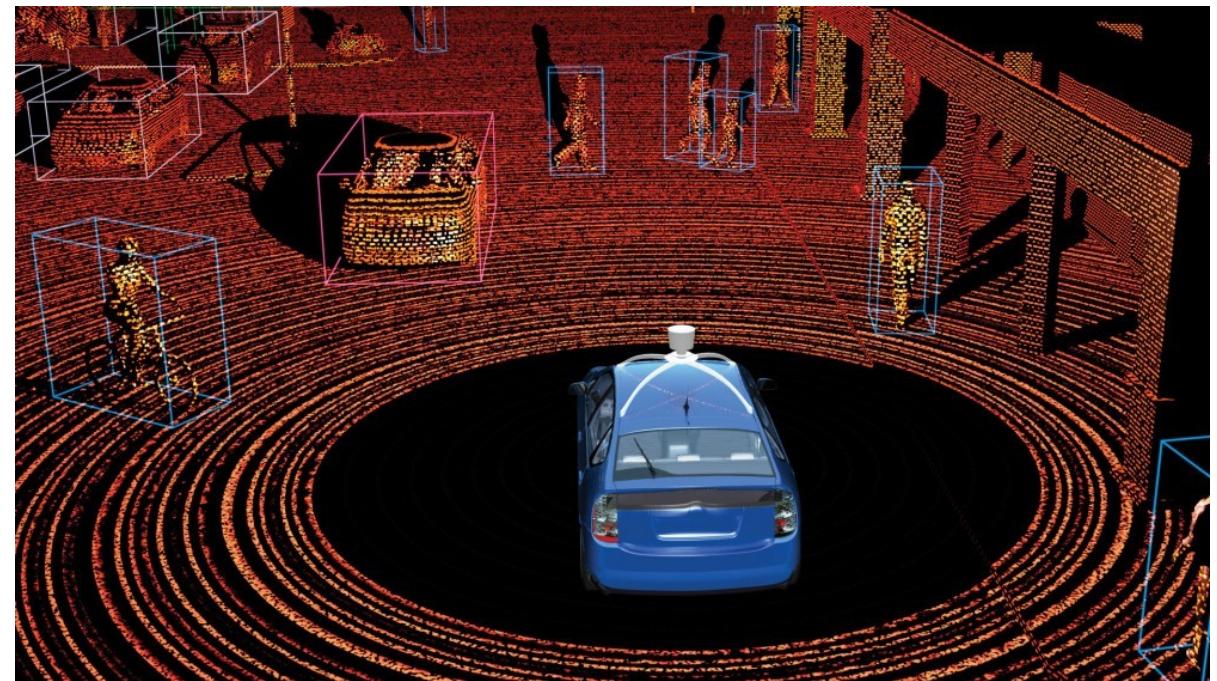
Why LiDAR is popular

- LiDAR is based on a laser emitting light pluses.
- The brightness level is higher than direct sunlight.
- It always get the front lighting direction.
- Direct Precise Distance Measurement.
- Less effected by the cloud condicions.
- Less post processing required.



LiDAR Challenges

- Low Resolution
- Outdoor performance
 - Dynamic range
 - Weather condition
- Expensive
- Light pollution
- Not possible
- in battle field



Camera based computer vision

- Lens choose:
 - Telescope
 - Fish eye
 - Normal
- Sensor size
 - SD 640x480
 - HD 1920x1080
 - 4K 3840x2160
- Special Cameras
 - Light field camera
 -



Breakthrough Technologies

- Computing Power – CPU, RAM, GPU
- Deep Learning
 - Neural Networks
 - Tensorflow, etc
- Sensors Fusion
 - LiDAR, RADAR,
 - Camera
- Libraries
 - Computer vision
 - Localization
 - Planning



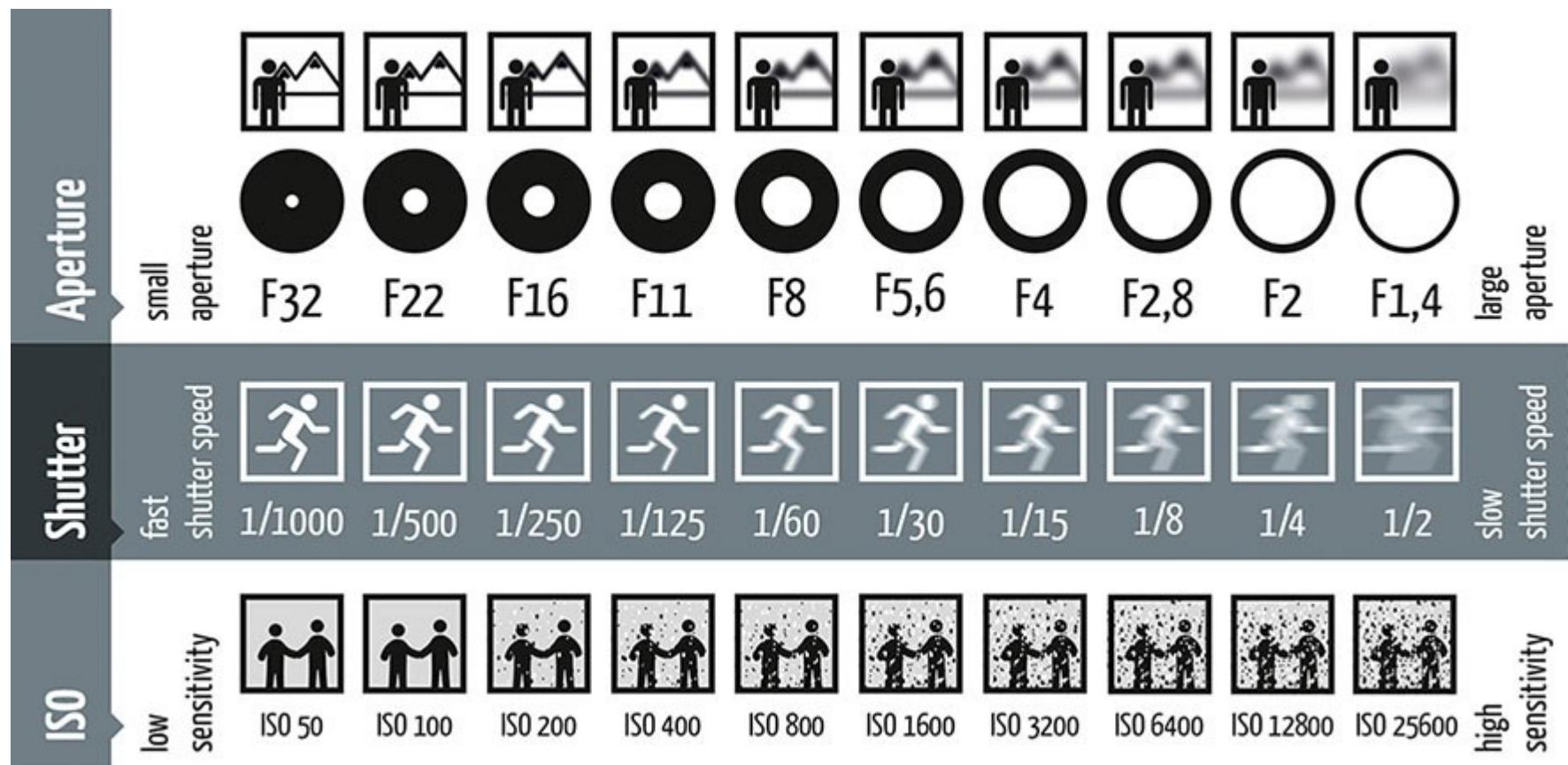
Camera based CV Challenges

- Very hard to have general solution
- Outdoor performance
 - Dynamic range
- Economic solution for indoor and industry
 - CCTV
 - Inspection
 - Quality Control



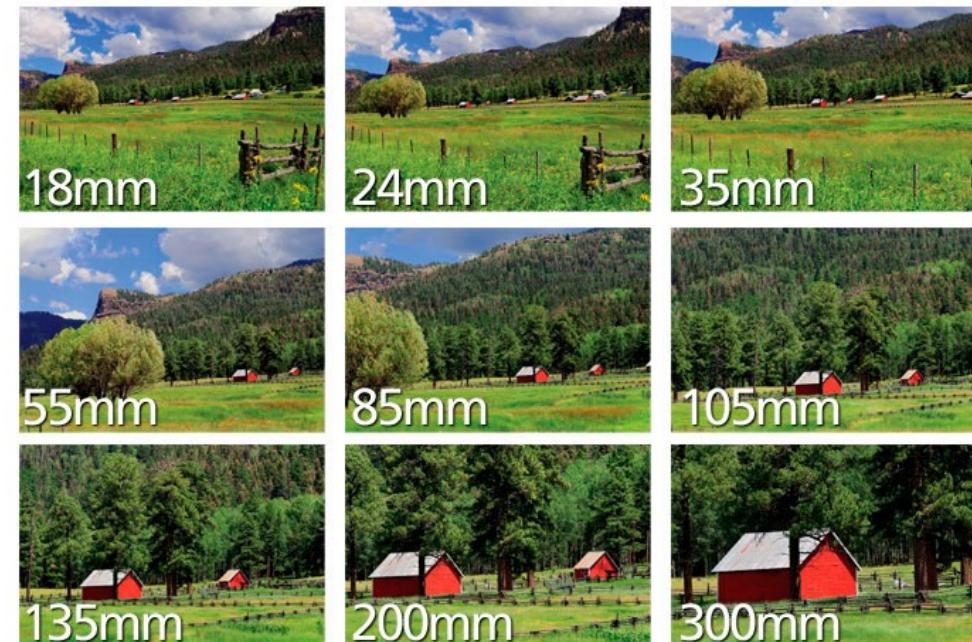
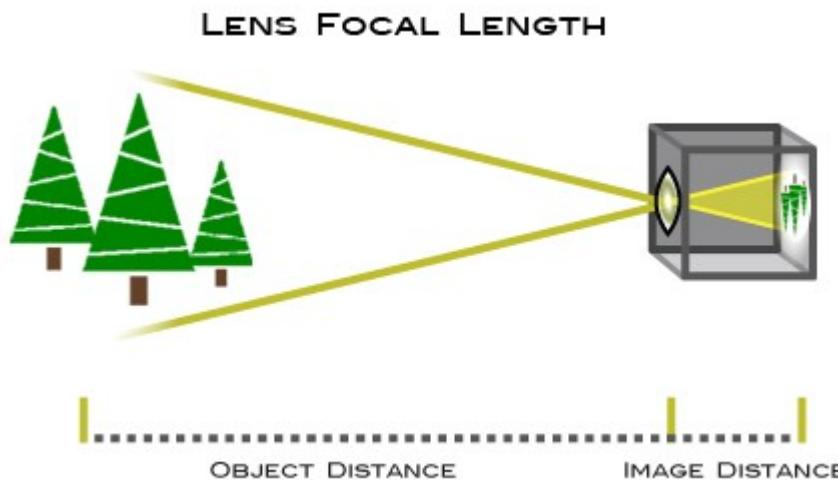
Lens: Shutter and Aperture

- Control the amount of light reaching the sensor
- More light result brighter image reading
-



Lens: Focal Length and Auto Focus

- Longer focal length, narrower angle of view
- Shorter focal length, wider angle of view
- 35mm -55mm close to human angle of view
- Assume all lens has auto focus, can always get clear picture



Lens: Distortion

- Barrel distortion
 - Occurrence in wide-angle lenses
- Pincushion distortion
 - Found in low-end telephoto lenses

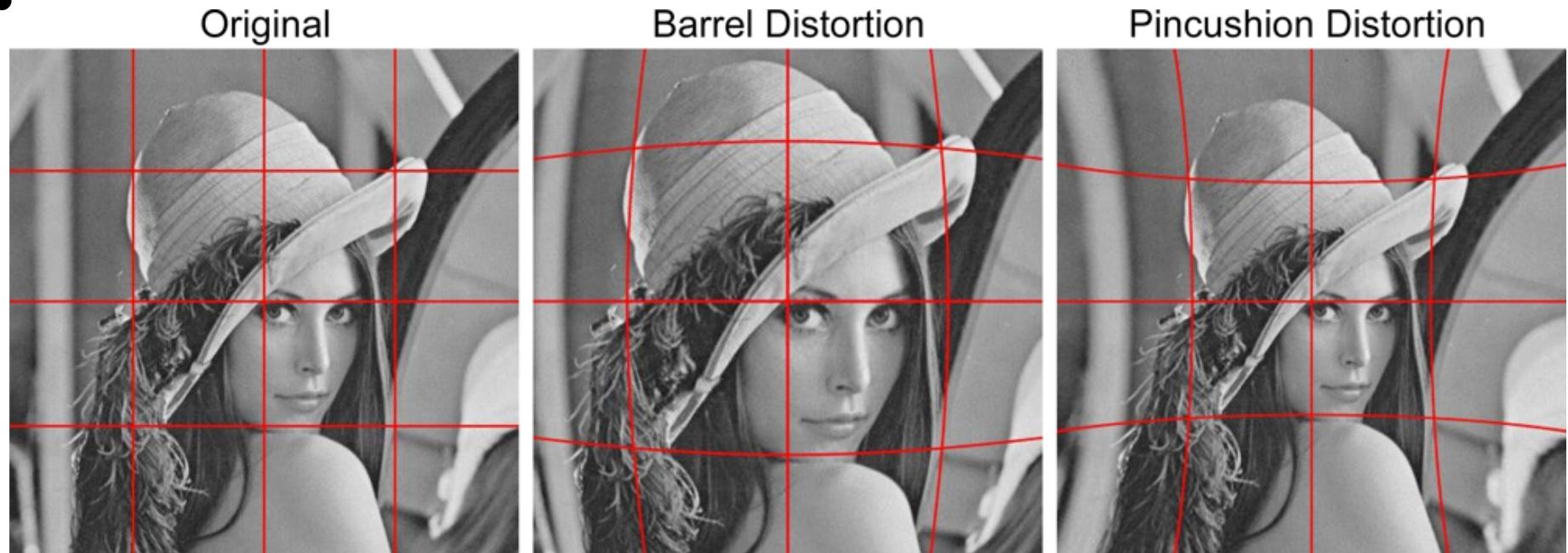


Image formation: Camera Body

- Don't need focus
 - Pinhole camera
 - Wide Angle
 -
- Focus before shot
 - Mirrorless camera
 - DSLR: Single
 - Lens Reflection
 -
- Focus after shot
 - Light Field camera
 -
 -
 -
 -



HOW PHASE DETECTION AUTOFOCUS WORKS

This cross-section of a Nikon D4 shows how the autofocus system works in most DSLRs

01 Main mirror

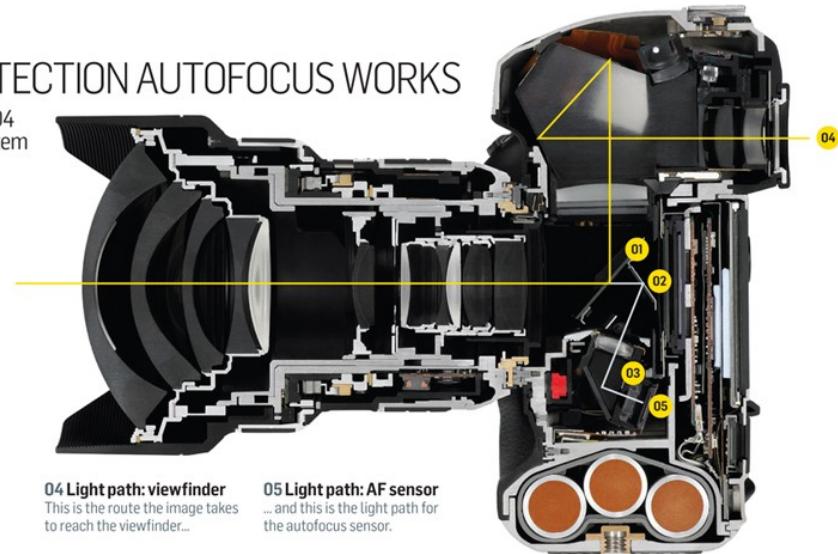
This is semi-silvered so that the image formed by the lens passes mostly up into the pentaprism and the viewfinder, but some passes through on to the sub-mirror behind.

02 Sub-mirror

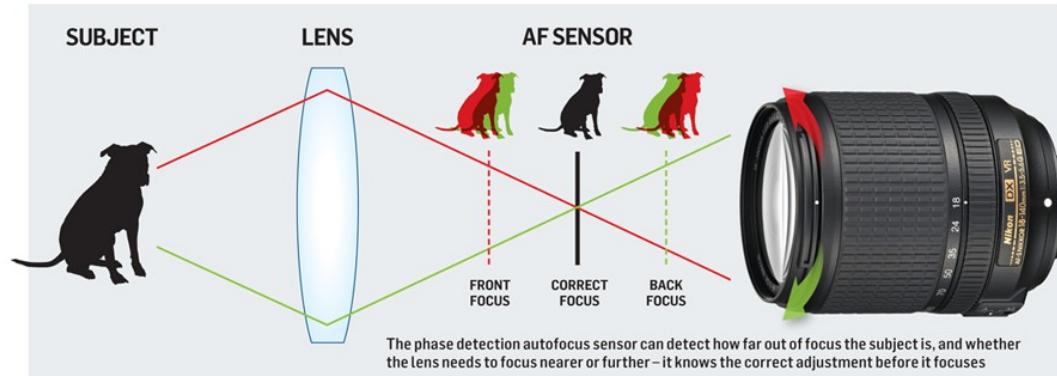
This reflects the image down into the base of the camera where the AF sensor is located. The sub-mirror flattens against the back of the main mirror when it flips up during the exposure.

03 Autofocus sensor

Different Nikon D-SLRs use different autofocus sensors – the pro-level D4 uses Nikon's top-of-the-range 51-point Multi-CAM 3500FX sensor.



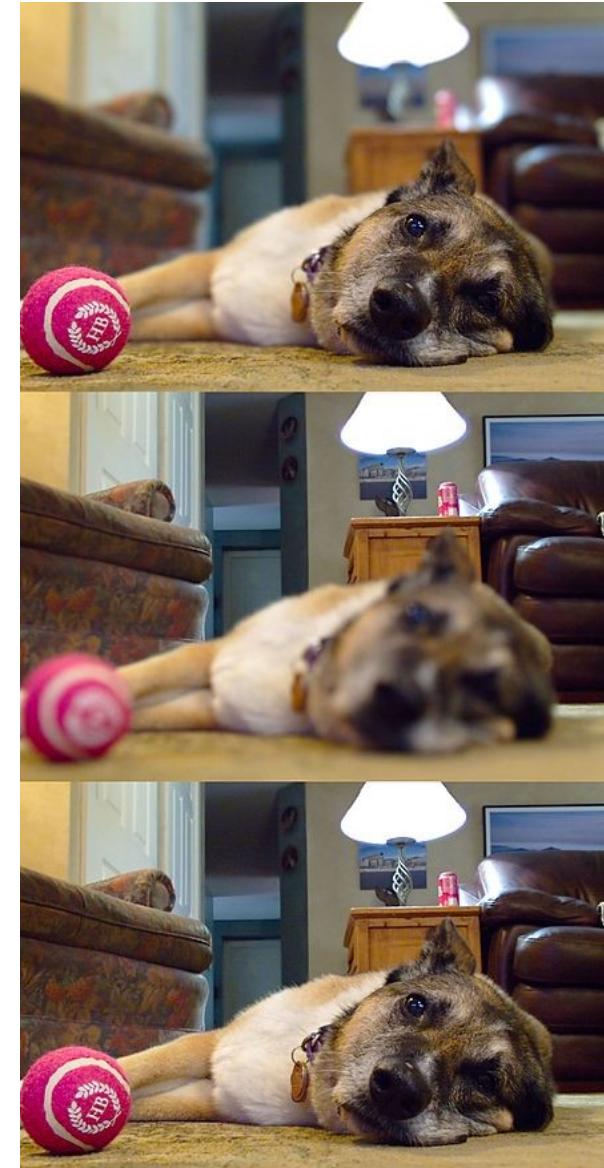
Unlike most other types of digital camera, digital SLRs use separate 'phase-detection' AF sensors



The phase detection autofocus sensor can detect how far out of focus the subject is, and whether the lens needs to focus nearer or further – it knows the correct adjustment before it focuses

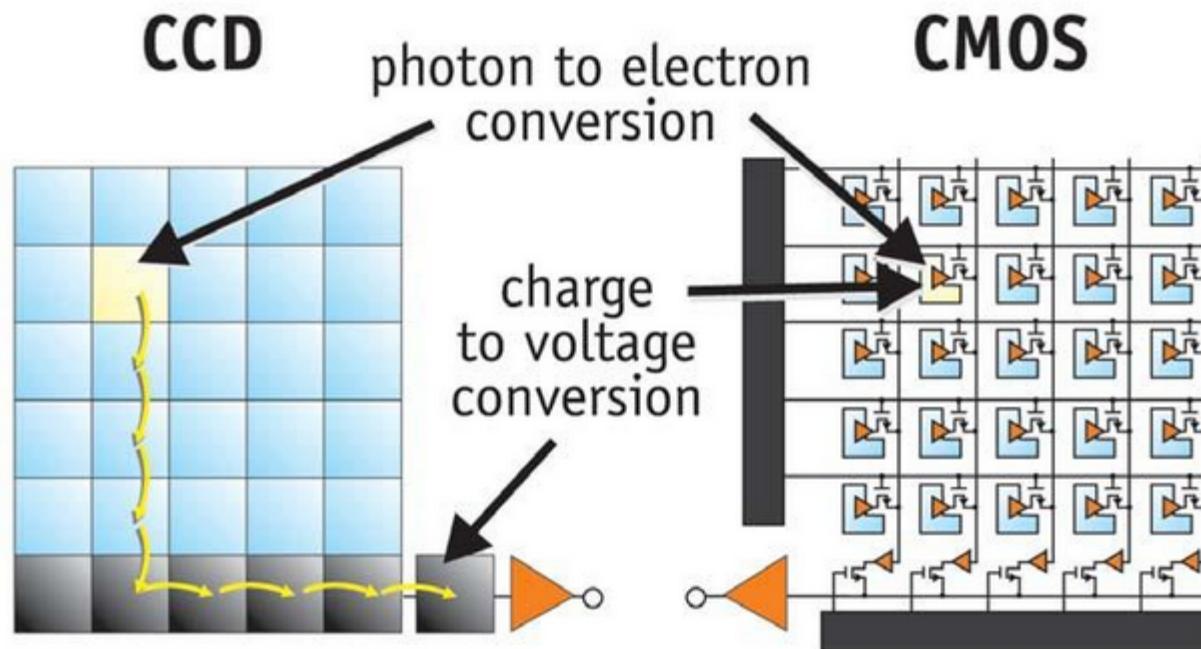
Light Field Camera

- Record at least 4 times more information for each shot
-
- Fast, no auto focusing required before the shot
-
- Low lighting requirement
-
- Depth of field and focus is produced by post processing
-
- 3D images
-
-

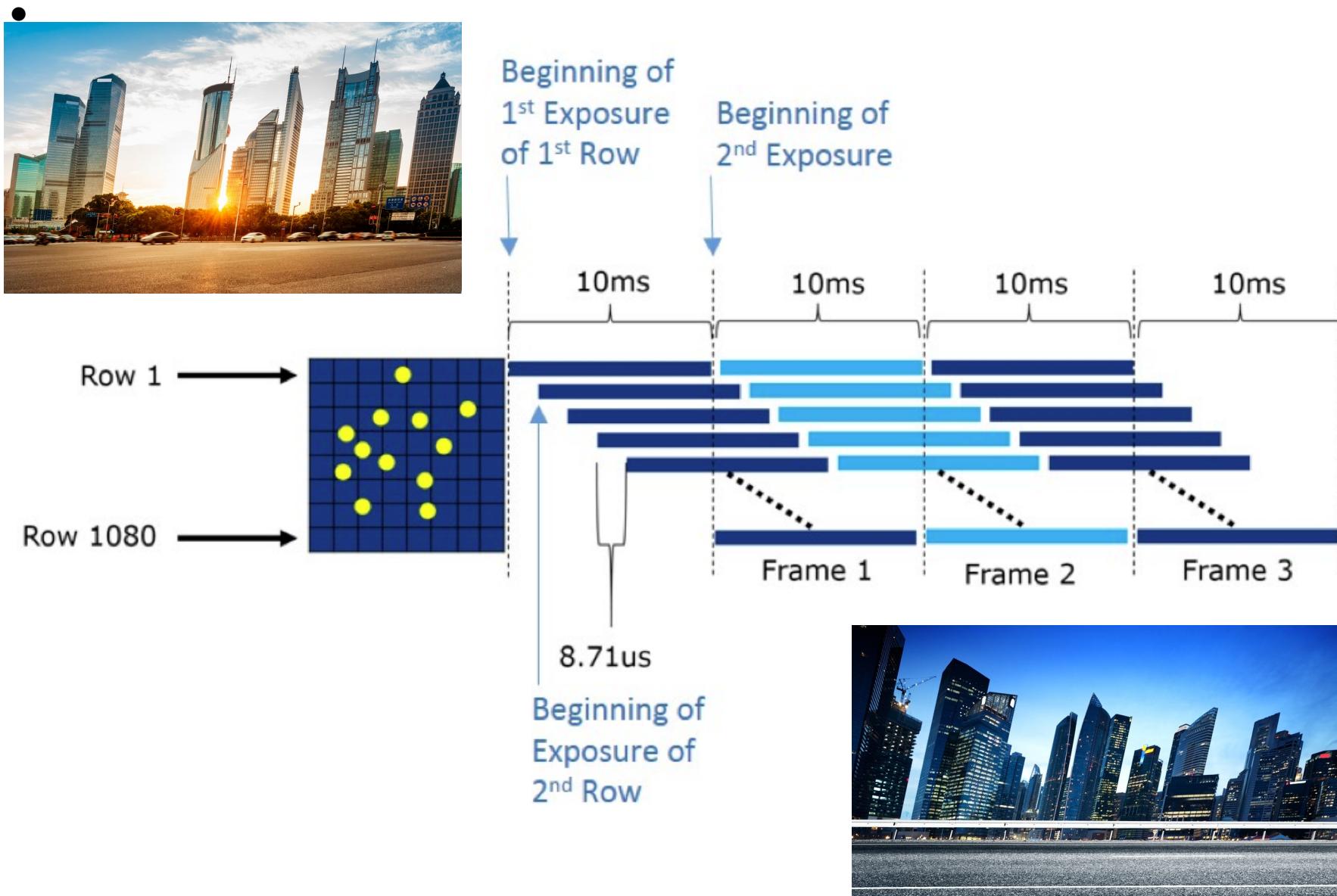


Camera Sensor

- CCD: Charge coupled device
 - Lower noise, expensive, global shutter
 - Higher power consumption
- CMOS: Complementary Metal Oxide on Silicon
 - Higher noise, cheaper, rolling shutter
 - Very Low power consumption
-



Global shutter vs Rolling shutter



Choose Camera Sensor

- Geometric reconstruction task:
 - The whole image is recorded at exactly the same time
 - Global shutter
- Film industry:
 - Global shutter
- Consumer Market
 - CMOS,
 - Rolling shutter
 - Lower power
 - consumtion
 - Lower cost
-



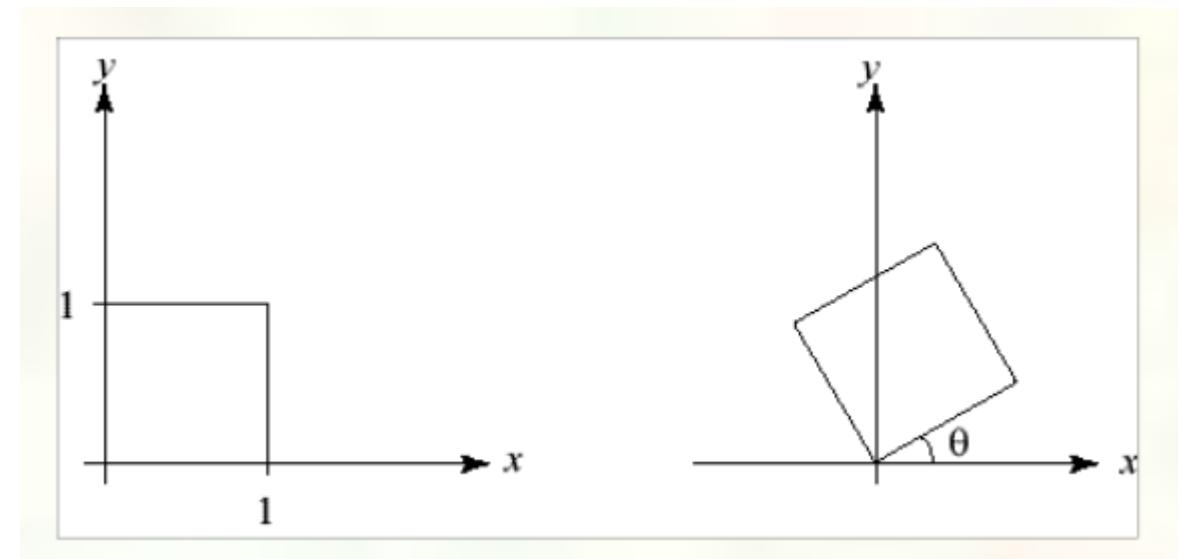
Digital Image Post Processing

- Enhancement
 - Redeye removal
 - Brightness
 - Sharpness
 - Dynamic Range
- Compress file size
- Filtering
- Editing
 - Crop
 - Resize



Transforms in 2D Computer Vision

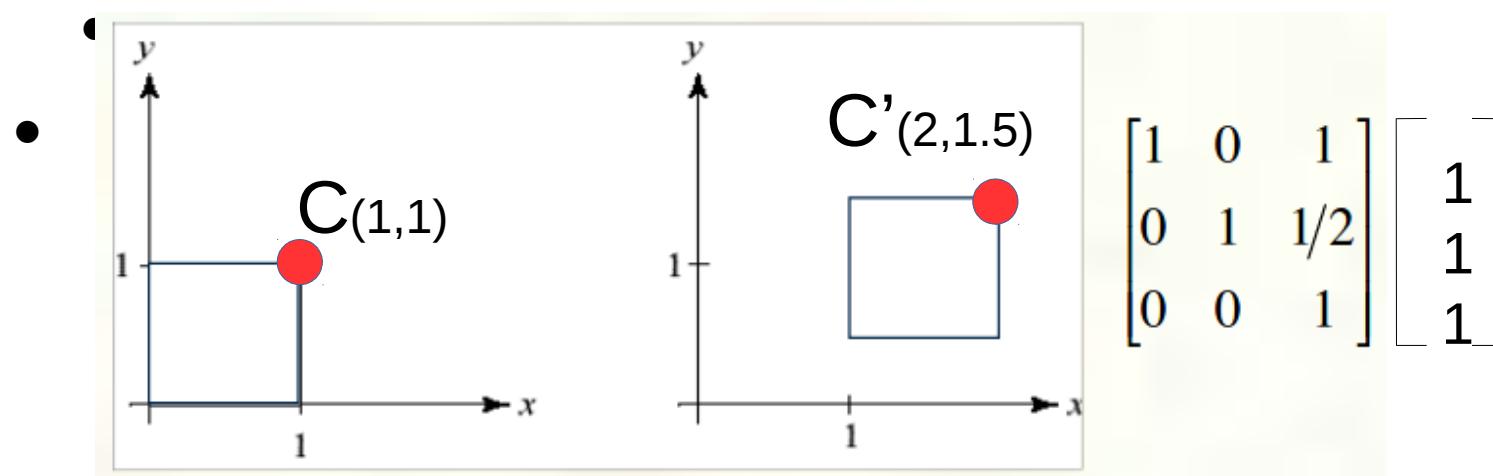
- Affine transforms
 - Translation
 - Rotation
 - Scaling
 - Reflection
 - Shear
-



- Transform between Coordinate systems
 - Local coordinate
 - Global coordinate

2D Transforms: Translation

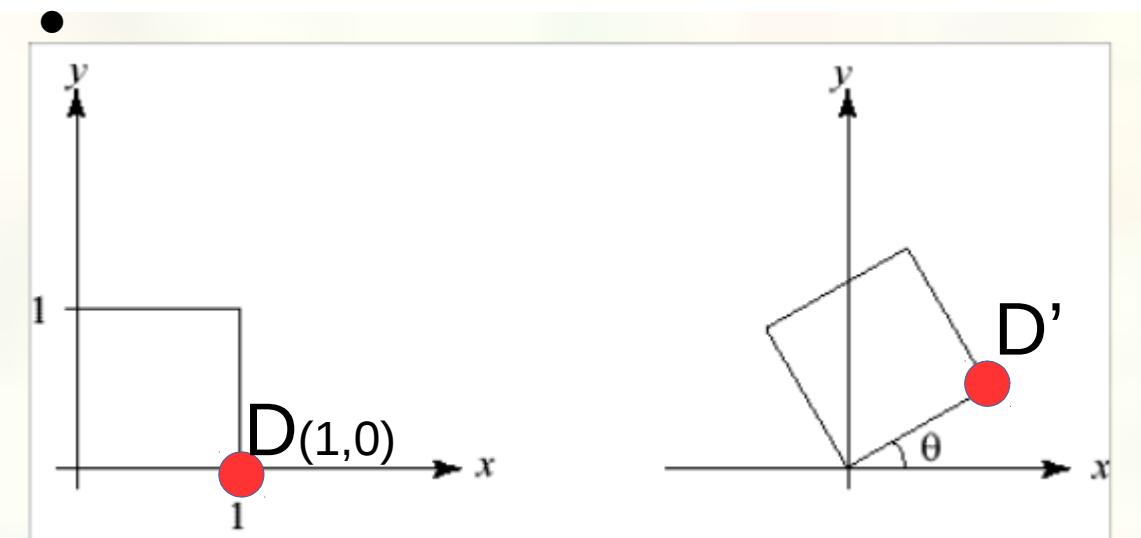
- Move a point $a = 1$ in x direction
- Move a point $b = 1/2$ in y direction
- Translation Matrix $T(a,b) =$



- Move a point $C(1,1)$ to $C'(2, 1.5)$
 - Matrix multiplication
 - Loop through all points if needed
 -

2D Transforms: Rotation

- Based on Origin (0,0),
- Counter-clock wise by angle theta
- Rotation Matrix $R(\theta) =$

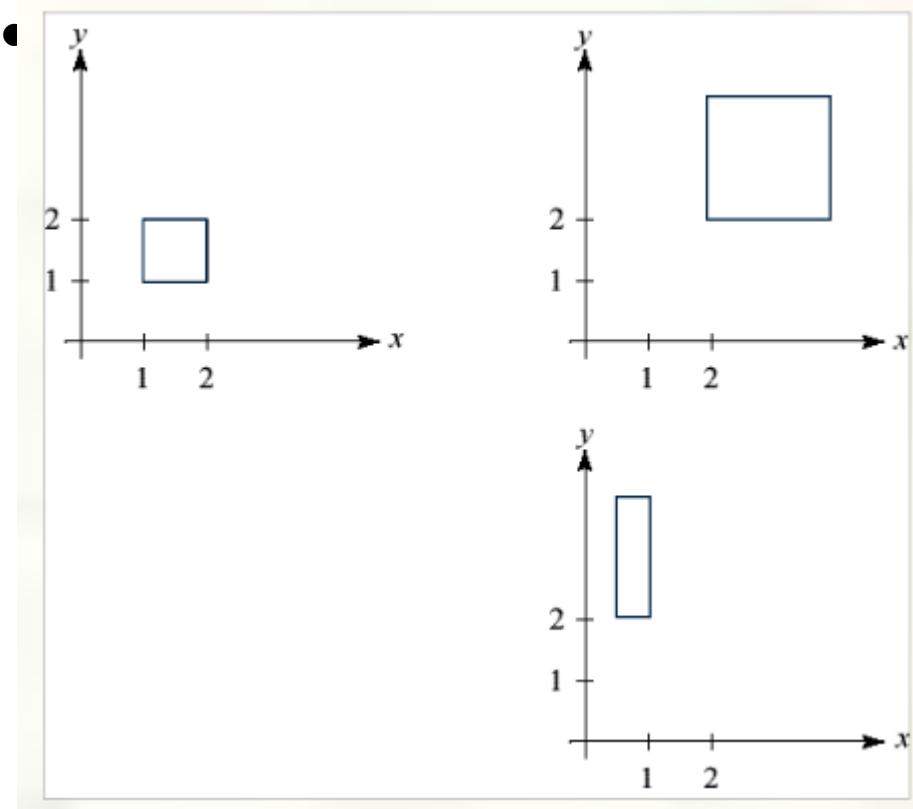


$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Move a point $D(1,0)$ to D'
 - Matrix multiplication
 - Loop through all points if needed
 -

2D Transforms: Scaling

- Scaling Matrix $S(s,t) = \begin{bmatrix} s, & 0, & 0 \\ 0, & t, & 0 \\ 0, & 0, & 1 \end{bmatrix}$
-



$$\begin{bmatrix} s, & 0, & 0 \\ 0, & t, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2, & 0, & 0 \\ 0, & 2, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.5, & 0, & 0 \\ 0, & 2, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

- Loop through all points if needed
-

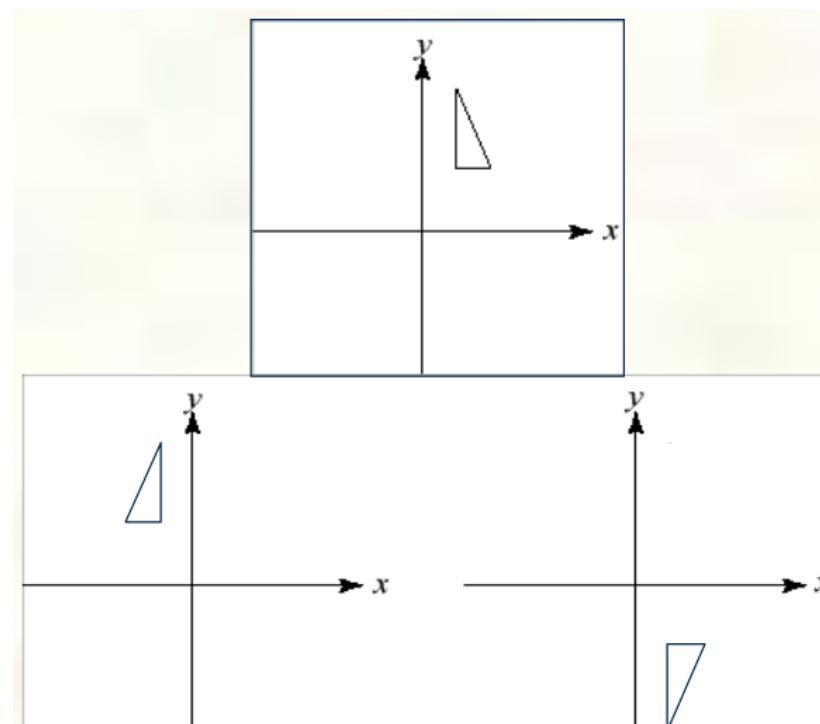
2D Transforms: Reflection

- Reflected about x or y axis Matrix

-

-

$$M(y) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



-

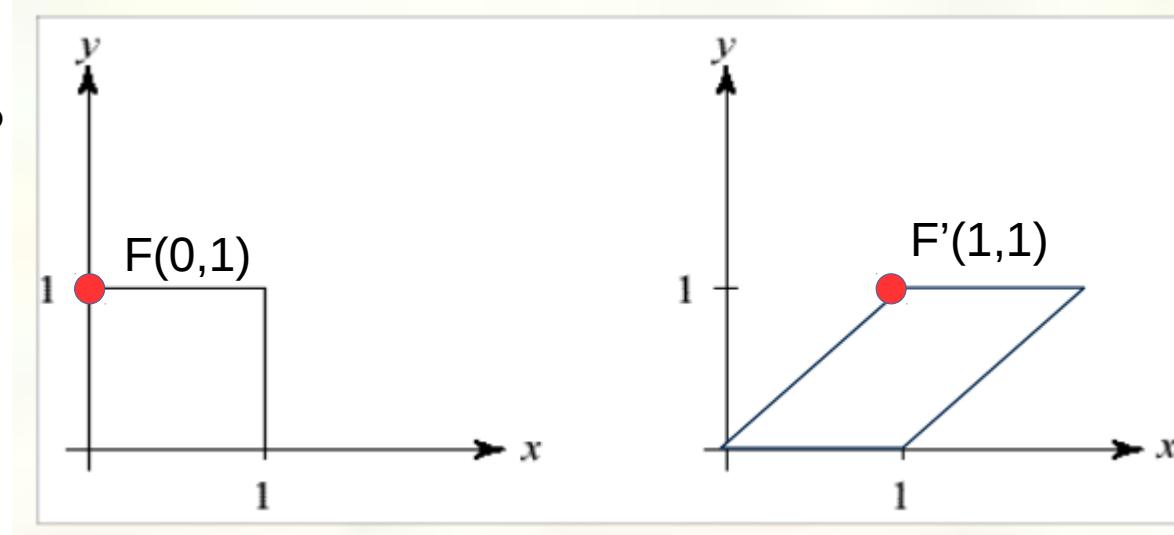
$$M(x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Loop through all points if needed

-

2D Transforms: Shear

- Use a to control x axis shear



$$S(x) = \begin{bmatrix} 1, & a, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

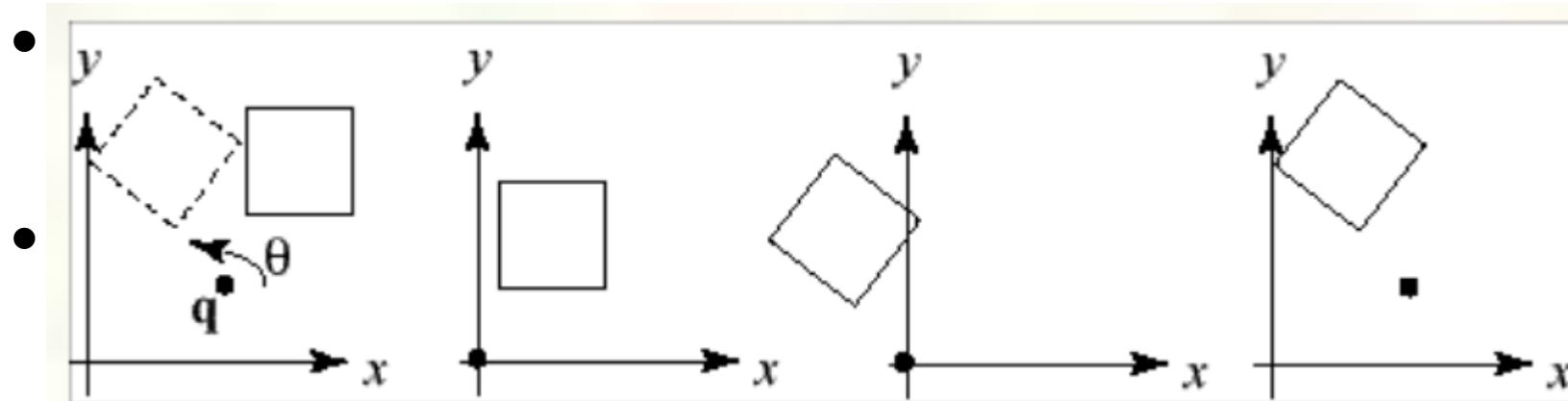
$a = 1$

$$S(y) = \begin{bmatrix} 1 & 0, & 0 \\ b, & 1, & 0 \\ 0, & 0, & 1 \end{bmatrix}$$

- Use b to control y axis shear
- Loop through all points if needed
-
-

Composition of Transforms

- Rotation about arbitrary points
- Given rotation point $q = [qx, qy, 1]^T$



- Translate q to origin
- Rotate
- Translate back
- $T(q)R(\theta)T(-q)q$
-
- Different multiplication order lead to different results
-

Local/Global Coordinates

- Global coordinate: GPS waypoints
 - Longitude: -114.185893
 - Latitude: 51.035545
- Local coordinate: Car Camera, LiDAR
 - Car always at origin(0,0)
 - All measure refer to car origin
 - Given car global location(x, y) and
 - heading direction (theta)
 - Convert GPS waypoints into local coordinate
- Now you can do all the magic!!!
 -
 -

Practical Computer Vision

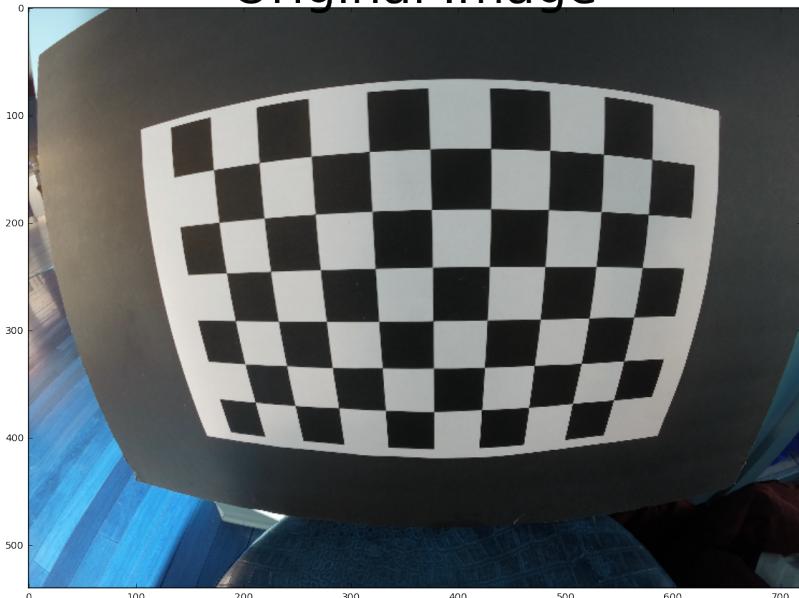
- Barrel Distortion Issues:
- Closer object appears larger than it should be
- Farther object appears even farther than it should be
- If you need use image to estimate distance and geometry, you need consider the distortion.



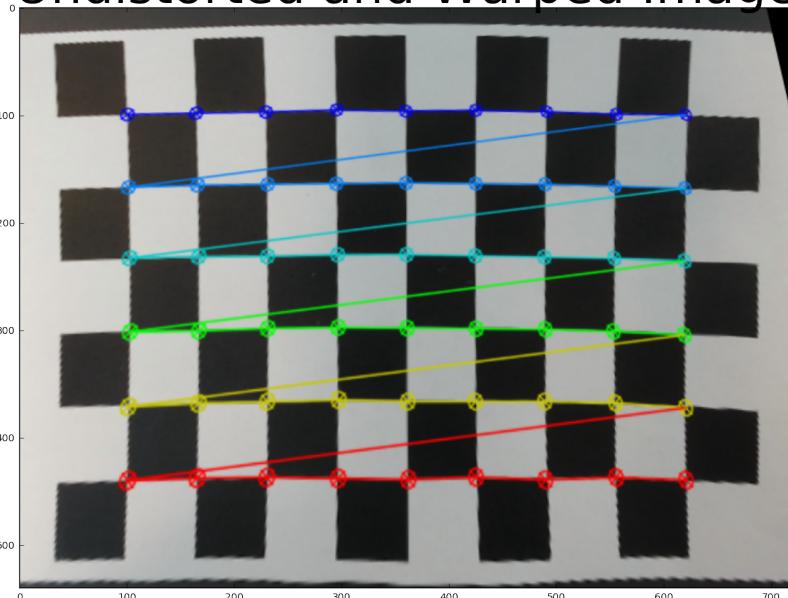
Camera Calibration: OpenCV tools

- cv2.findChessboardCorners()
- cv2.calibrateCamera()
- cv2.undistort()
- cv2.warpPerspective()

Original Image

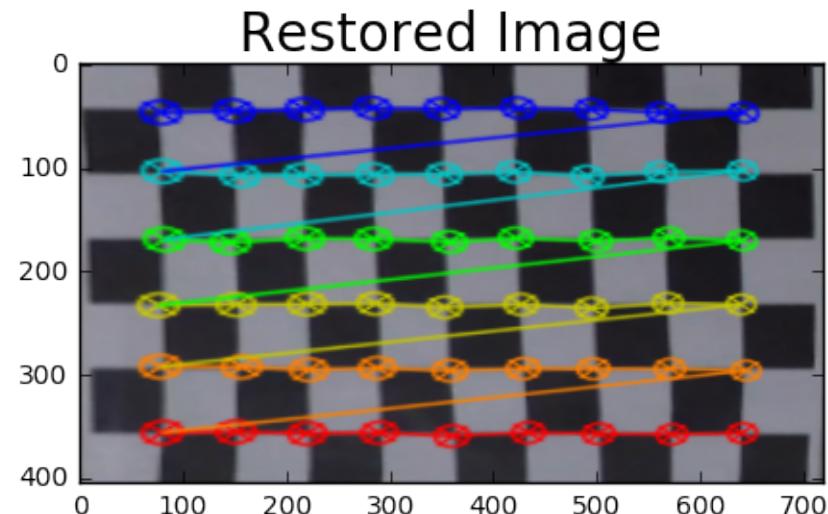
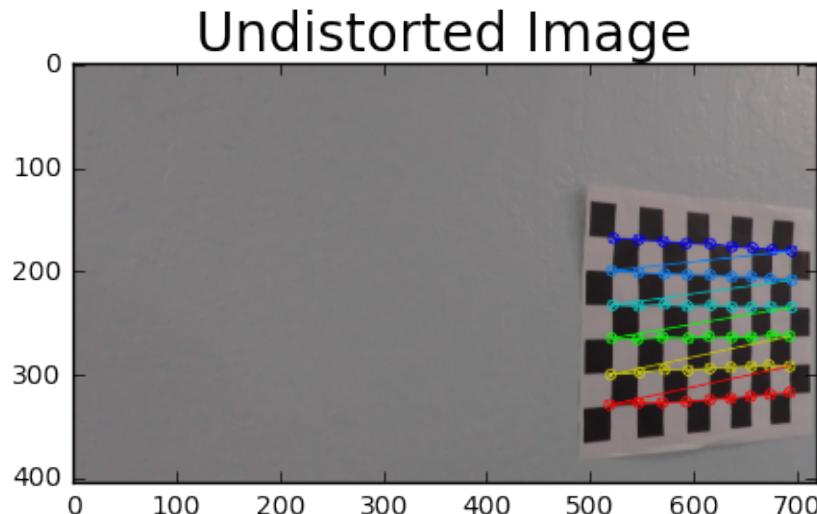


Undistorted and Warped Image



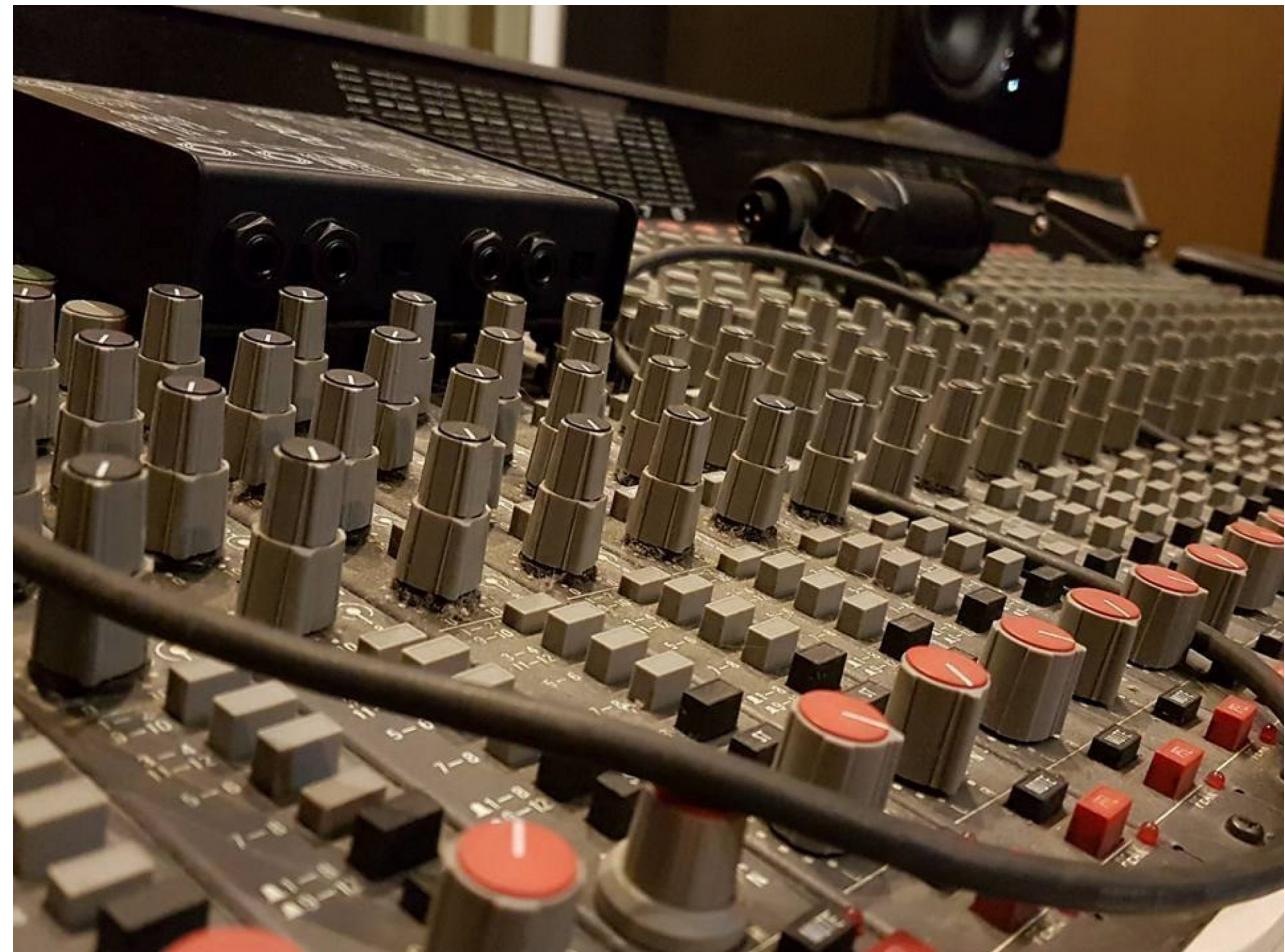
Camera Calibration: OpenCV Tips

- Take 20-30 pictures for calibration to cover all image areas and angles
- Each resolution has its own matrix
- Save the matrix for future use



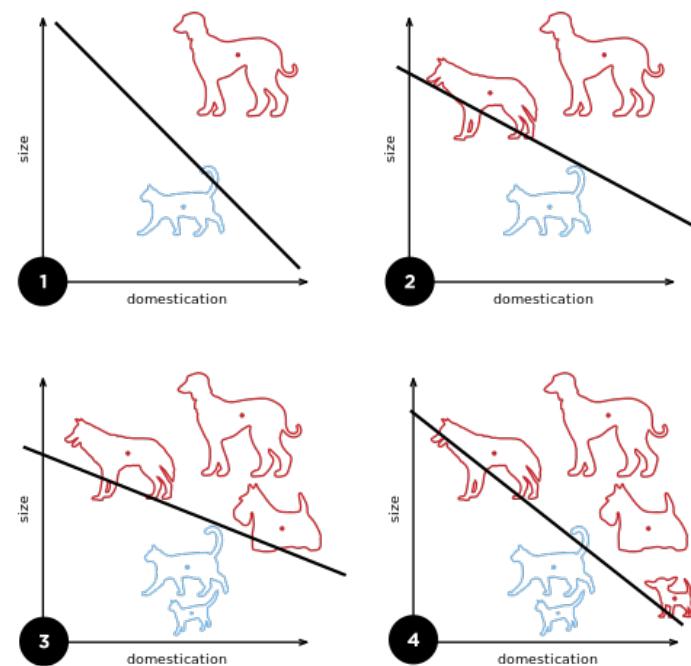
Traditional CV vs Deep Learning CV

- How to tune the filter parameters
- Speed
- Training
- Power



Early Artificial Neural Networks

- 1986 Geoffrey Hinton found the way to train the multi-layer neural network
 - Backpropagation(Backward propagation of errors)



Brutal Force

KEY MOMENTS IN DEEP-LEARNING HISTORY 1989-1997

1989

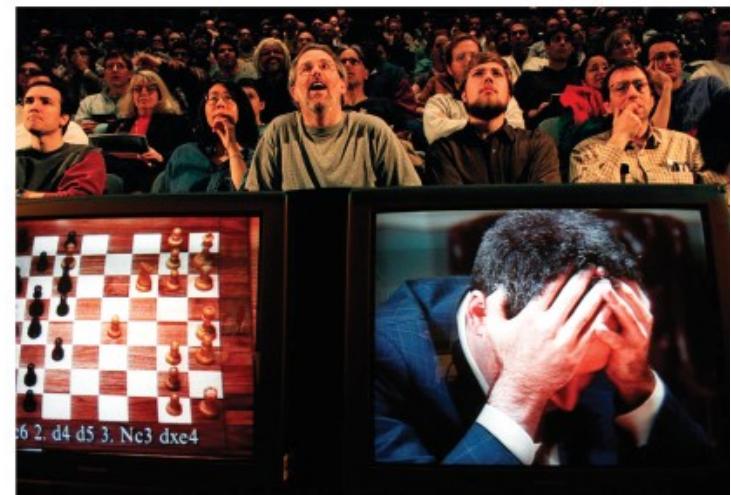
French researcher Yann LeCun, then at Bell Labs, begins foundational work on a type of neural net that becomes crucial for image recognition.

1991

German researchers Sepp Hochreiter and Jürgen Schmidhuber pioneer a neural net with memory features, which eventually proves superior for natural-language processing.

1997

IBM's Deep Blue beats **world champion Garry Kasparov** (right) in chess using traditional AI techniques.



STAN HONDA—AFP/GETTY IMAGES

Dataset

KEY MOMENTS IN DEEP-LEARNING HISTORY 1990's-2011

Mid-1990s

Neural nets fall into disfavor again, eclipsed by other machine-learning techniques.

2007

Fei-Fei Li founds ImageNet and begins assembling a database of 14 million labeled images that can be used for machine-learning research. →



2011

Microsoft introduces neural nets into its speech-recognition features.

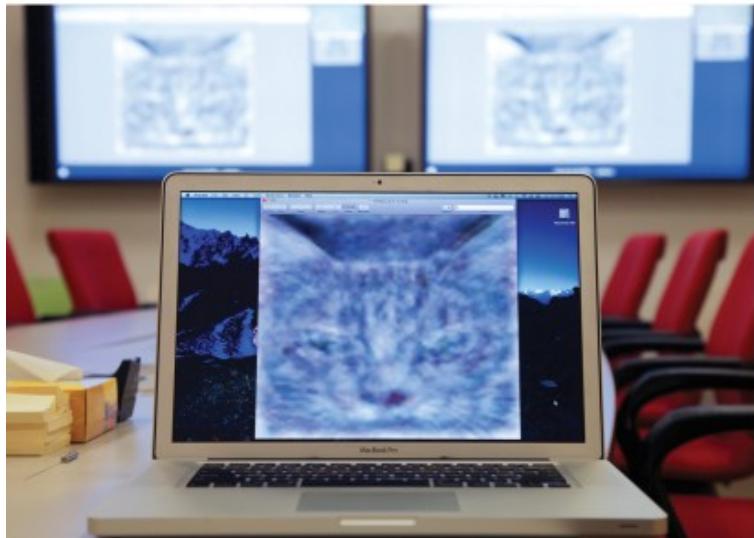
2011

IBM's Watson beats two champions at Jeopardy using traditional AI techniques.

CARLOS CHAVARRIA—THE NEW YORK TIMES/REDUX PICTURES

Object recognizing

KEY MOMENTS IN DEEP-LEARNING HISTORY 2012-2013



2012
JUNE
Google Brain publishes the “cat experiment.” A neural net, shown 10 million unlabeled YouTube images, has trained itself to recognize cats.

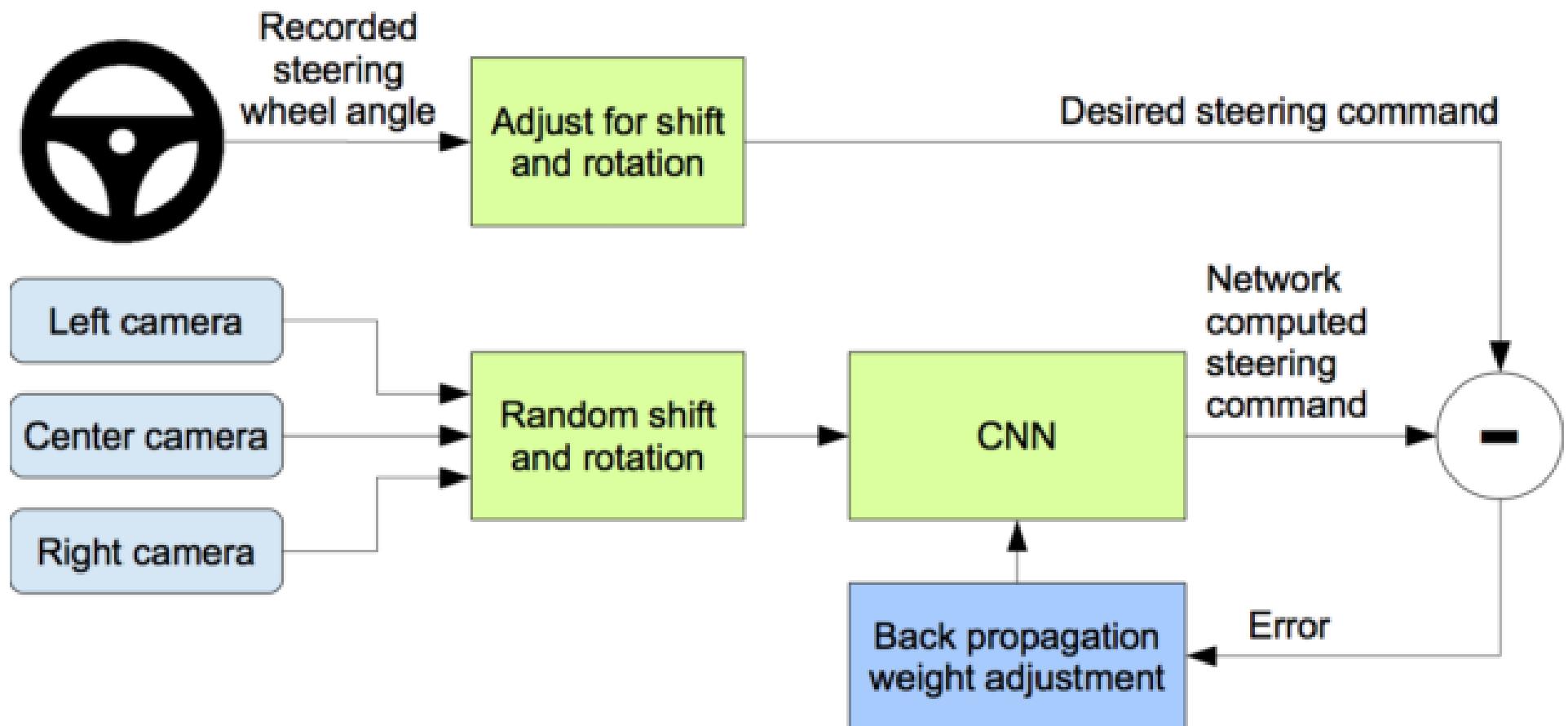


AUGUST
Google introduces neural nets into its speech-recognition features.
OCTOBER
A neural net designed by two of Hinton’s students wins the annual ImageNet contest by a wide margin.

2013
MAY
Google improves photo search using neural nets.

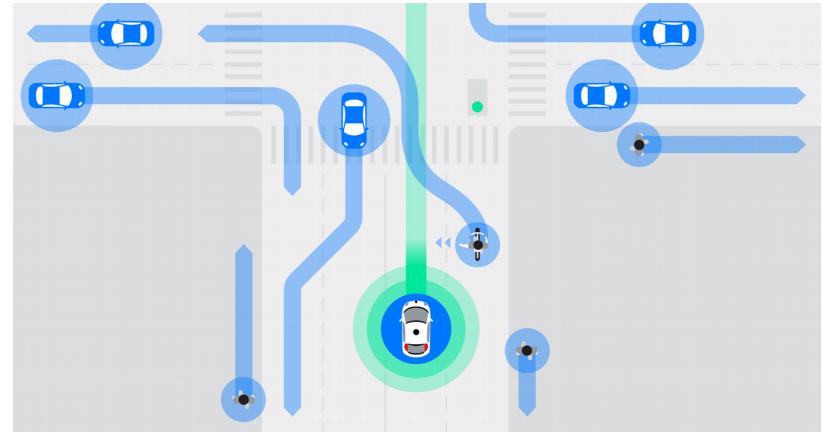
JIM WILSON—THE NEW YORK TIMES/REDUX PICTURES

End to End Learning



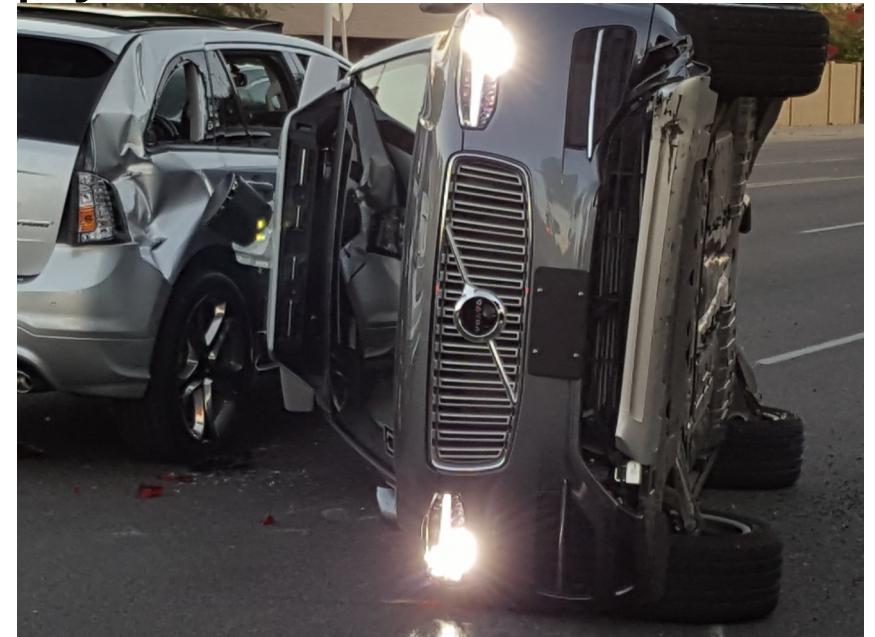
Environment

- In real situation:
- City/Highway/Traffic/Construction/Emergency
- Weather/Rain/Snow/Day/Night

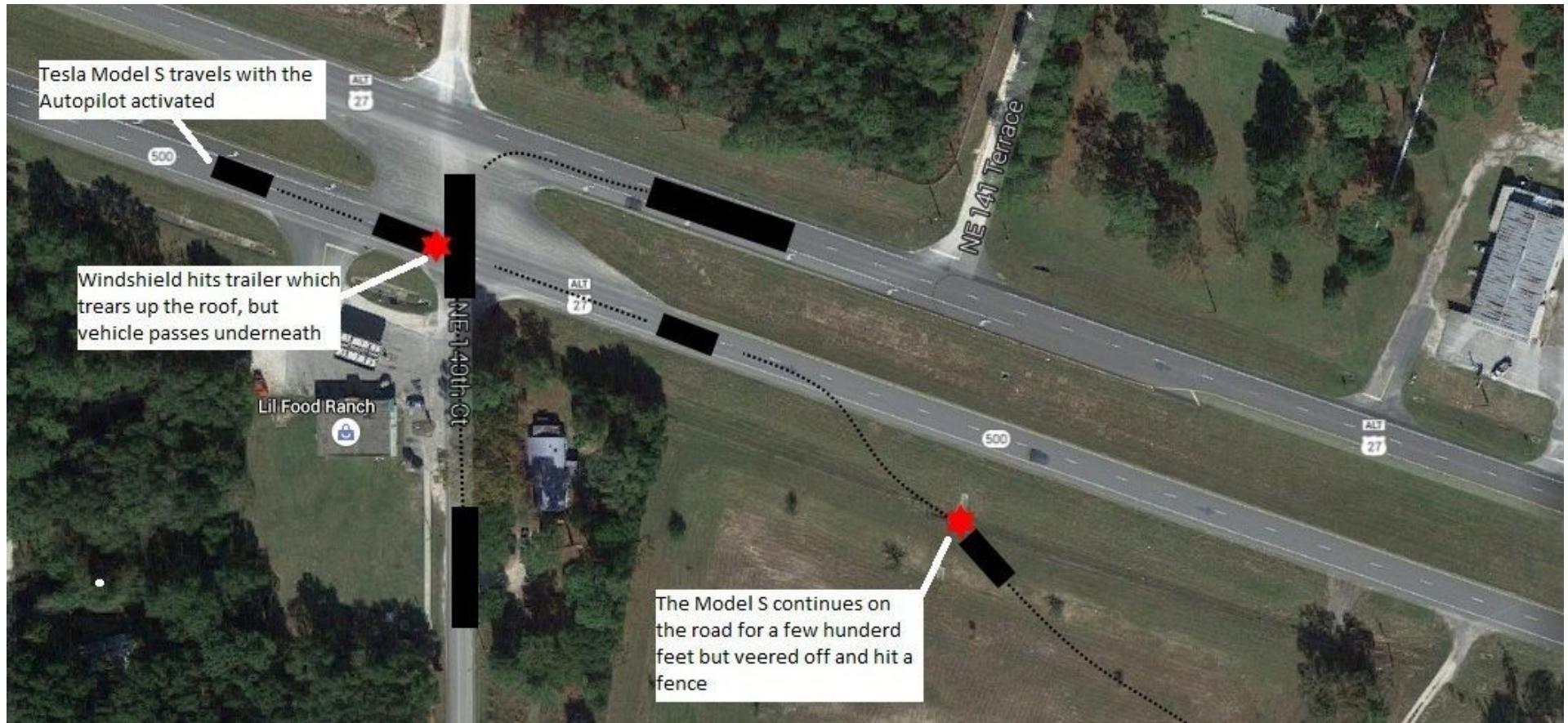


Compansate/Fall Safe

- System Failures:
 - Hardware failure 10%
 - Software bugs 70%
 - Unexpected event 20% / not able to generalize the machine learning and apply to new events.
- Plan B
 - Redundancy
 - Fall safe
 - Reboot
- Call for help



Tesla/Semi Fatal Crash



Can you see the truck



- Tesla can't while the accident happened

SAE Autonomy Levels

	Name	Definition	Execution	Monitoring	Fallback	System Capability
0	No Automation		Human	Human	Human	N/A
1	Drive Assistance	Hands on	Human /system	Human	Human	ACC, LKA, Parking
2	Partial Automation	Hands off (mandatory check in)	System	Human	Human	Testing stage
3	Conditional Automation	Eyes off	System	System	Human (traffic jam)	Stop and go
4	High Automation	Mind off In optimital areas (geofencing, traffic jam, parking lot etc)	System	System	System (abort the trip if driver does not retake controls)	Robot shuttle bus
5	Full Automation	Steering Wheel optional	System	System	System	Robot taxi

Technical Challenges

- Computer vision
- Sensor Fusion
- Localization
- Path Planning
- Controller



The screenshot shows a simulation interface for a self-driving car. On the left, a terminal window displays log data from the "self_driving_car_nanodegree_program". The data includes global coordinates (globalx, globaly) and pixel coordinates (ptx, pty) for various objects. On the right, the simulation interface shows a road with other cars. A yellow car is highlighted with a green trajectory line. Red text on the screen reads "Max Acceleration Exceeded!", "Collision!", and "Violated Speed Limit!". The interface also displays performance metrics: Distance Without Incident (Best: 0.05 Miles, Curr: 0.00 Miles), Timer (0:00:00), AccT (-31 m/s^2), AccN (2 m/s^2), AccTotal (31 m/s^2), and Jerk (4 m/s^3). A speedometer icon shows 66.35 MPH. At the bottom, a terminal window shows code related to path planning, including comments about calculating smoother paths and distances.

```
self_driving_car_nanodegree_program
globalwx: 2309.49 globalwy: 2409.94
globalwx: 2317.37 globalwy: 2444.34
globalwx: 2322.2 globalwy: 2469.31
globalwx: 2328.32 globalwy: 2583.4
globalwx: 2334.39 globalwy: 2697.24
globalwx: 2337.79 globalwy: 2570.78
globalwx: 2338.1 globalwy: 2587.07
globalwx: 2337.8 globalwy: 2611.5
globalwx: 2338.10 globalwy: 2656.66
globalwx: 2338.1 globalwy: 2658.92
globalwx: 2338.9 globalwy: 2682.89
globalwx: 2339.1 globalwy: 2709.25
globalwx: 2339.10 globalwy: 2723.0
globalwx: 2339.2 globalwy: 2739.69
globalwx: 2339.39 globalwy: 2773.33
globalwx: 2336.41 globalwy: 2799.05
ptx: 2298.8 pty: 2341.32
ptx: 2298.8 pty: 2344.4
ptx: 2284.23 pty: 2345.87
ptx: 2293.61 pty: 2374.36
ptx: 2302.97 pty: 2402.87
[10,2269.111,2277.384,2.891631,13.68515,2354.795,5.834941], [1,2264.772,2276.76
,4.443,2252.896,2.788815,19.14973,2329
,7.14.38221,2385.596,16.01379],[4,2278
,048375],[5,2278.528,2322.019,4.795273
,1.2278.528,2322.019,4.795273],[6,2278
,1.2441.711,5.320045],[8,2251.044,218
,9,2274.858,2284.926,3.2199,15.41058
,902,16.62994,2469.546,1.934325],[11,7
,1.2558 deg car_yaw in rad: 1.33092
]
*/ cout << "dist_inc :" << dist_inc << "ref velocity: " << ref_vel<< endl;
// calculate the smoother path
double localxx = 0;
double localxy = 0;
for(int i=0; i<10; i++){
    ly[i] = smooth_lanes(localxx, localxy, lx[i], ly[i]);
    double dist = distance(localxx, localxy, lx[i], ly[i]);
    if (dist > speed || dist < speed*0.8){
        double heading = atan2(lx[i]-localxx, ly[i]-localxy);
        lx[i] = localxx + smooth_speed(double(i))*cos(heading);
        ly[i] = smooth_lanes(lx[i]);
        dist = distance(localxx, localxy, lx[i], ly[i]);
    }
    localxx = lx[i];
}
```



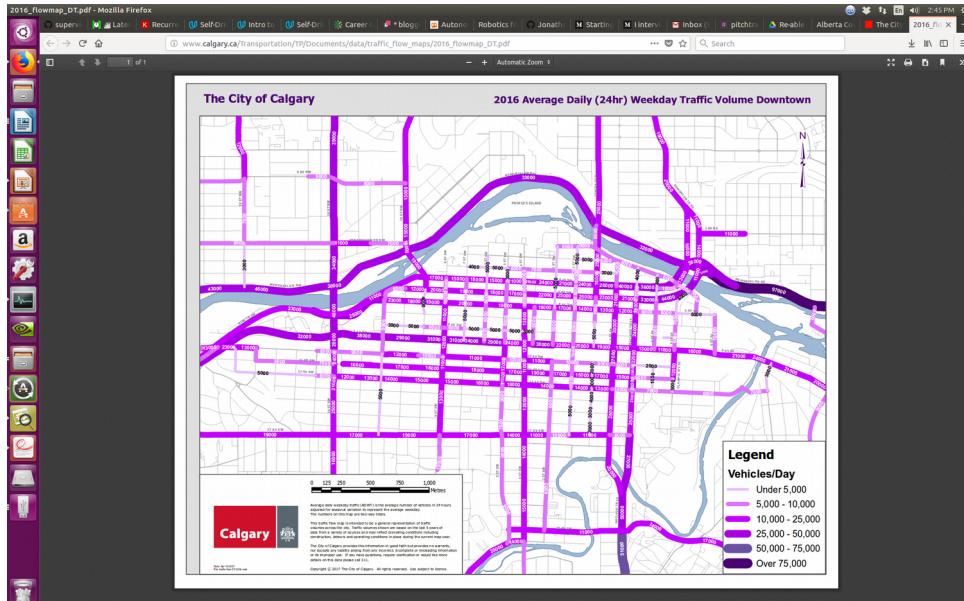
Solutions - Perception

- Traditional Computer vision
- Deep learning object detection and recognition
 - Traffic signs
 - Traffic lights
 - Objects



Our Focused Area

- Neural Network design and test
- Create training and validation dataset
- Produce control units
- User experience pilot project



Building Self Driving Car - Local Dataset - Night

Current Plan

- Launch Self-Driving User experienment platform prototype 1 and 2 early 2018
- Approach Industrial leader for private testing
- Approach Provincial government and City for public testing permit
- Recent work:
 - <https://www.youtube.com/watch?v=7nwrwjLZxcw>
 -

Team Leader Profile

- Calvenn Tsuu P. Eng PMP
 - Self-Driving Car Engineer
 - Mechanical Engineer
-
- Experiences in Robotics, computer vision, Neural Network, CNC machine tools, Automation and Project Management
<https://www.linkedin.com/in/calvenn-tsuu-94a90317/>
 - Phone: 403-926-5689
 - Email: calvenn.tsuu@gmail.com